DOCUMENT RESUME

ED 389 716 TM 024 196

AUTHOR	Schmitt, Alicia P.; And Others
TITLE	Differential Speededness and Item Omit Patterns on the SAT.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-91-50
PUB DATE	Aug 91
NOTE	49p.; A synopsis of this paper was presented at the Meeting of the National Council on Measurement in Education (Boston, MA, April 1990).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Asian Americans; Black Students; *College Entrance Examinations; *Difficulty Level; Ethnic Groups; Females; High Schools; *High School Students; Hispanic Americans; *Item Bias; Males; Mathematics Tests; *Racial Differences; Responses; Scores; Verbal Tests
IDENTIFIERS	*Omitted Responses; Scholastic Aptitude Test; *Speededness (Tests)

ABSTRACT

Two editions of the Verbal and Mathematical portions of the Scholastic Aptitude Test (SAT) were used to study differential speededness and differential omission and the relationships among differential item functioning (DIF), differential omission, and item difficulty for Asian Americans, Blacks, Hispanics, and females in a total sample exceeding 300,000 SAT examinees in 1984 and 1985. Consistent and replicable evidence of differential speededness was found for Blacks and Hispanics. Use of an unspeeded criterion for matching in place of the traditional total score, which contains speeded items, did not affect the DIF analyses of speeded items. A strong artifactual negative relationship between DIF and differential omission was found. The relationship between differential omission and difficulty was consistently positive on the Verbal sections for all comparison groups except the Asian-American group, for whom it was consistently negative. On the mathematical sections, this relationship was consistently found for the female/male comparison, for whom it was negative. Finally, the relationship between difficulty and DIF was negative but smaller than previously observed. (Contains 8 tables, 8 figures, and 31 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-91-50

ED 389 716

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Differential Speededness and Item Omit Patterns on the SAT

Alicia P. Schmitt
Neil J. Dorans
Carolyn R. Crone
Behroz T. Maneckshana

Educational Testing Service
Princeton, New Jersey
August 1991

BEST COPY AVAILABLE

TM024196

Differential Speededness and Item Omit Patterns on the SAT

Alicia P. Schmitt[1,2,3], Neil J. Dorans,
Carolyn R. Crone, and Behroz T. Maneckshana

Educational Testing Service

Running head: DIFFERENTIAL SPEEDEDNESS AND OMIT

---

1 A synopsis of this paper was presented at the National Council on Measurement in Education, Boston, April, 1990.

2 Funding for this research was provided by the Admissions Testing Program of the College Board and Educational Testing Service. The interpretation of results represent the opinions of the authors and should not be viewed as statements of ETS or the College Board.

3 The authors thank Edward Kulick and Michael Zieky for their reviews and comments on an earlier version of the paper, and Elise Sharrett for her assistance in revising the paper.

Copyright © 1991. Educational Testing Service. All rights reserved.

Abstract

Two editions of the Verbal and Mathematical portions of the Scholastic Aptitude Test (SAT) were used to study differential speededness and differential omission and the relationships among differential item functioning (DIF), differential omission, and item difficulty for Asian-Americans, Blacks, Hispanics, and females. Consistent and replicable evidence of differential speededness was found for Blacks and Hispanics. Use of an unspeeded criterion for matching in place of the traditional total score, which contains speeded items, does not affect the DIF analyses of the speeded items. A strong artifactual negative relationship between DIF and differential omission was found. The relationship between differential omission and difficulty was consistently positive on the Verbal sections for all comparison groups except the Asian-American group, for whom it was consistently negative. On the Mathematical sections, this relationship was only consistently found for the female/male comparison, for whom it was negative. Finally, the relationship between difficulty and DIF was negative but smaller than previously observed.

Differential Speededness and Item Omit Patterns on the SAT

Differential item functioning (DIF), under one rubric or another, has been an ongoing consideration by psychometricians and educators for nearly forty years. Item level differences that could result in incorrect estimation of abilities for examinees from different subpopulations have been of particular concern during the second half of the last decade. Because of this special emphasis, numerous papers have been published and or presented about methods to estimate DIF (Dorans & Kulick, 1986; Green, Crone, & Folk, 1989; Holland & Thayer, 1988; Scheuneman & Bleistein, 1989; Shepard, Camilli, & Williams, 1985), about sample dependent sources of DIF (Bleistein & Wright, 1987; Chen & Henning, 1985; Kulick & Dorans, 1983) and about item factors related to DIF (Alderman & Holland, 1981; Bleistein, Schmitt, & Curley, 1990; Breland, Stocking, Pinchak, & Abrams, 1976; Scheuneman, 1987; Schmitt, 1985, 1988; Schmitt & Bleistein, 1987; Schmitt, Curley, Bleistein, & Dorans, 1988; Schmitt & Dorans, 1990, in press). In addition, examinee response style factors related to DIF have also been addressed. These are differential speededness (Dorans, Schmitt, & Curley, 1988; Schmitt & Bleistein, 1987) and differential omission (Rivera & Schmitt, 1988). Item difficulty has also been found to be related to DIF and interrelated to other DIF factors (Kulick & Hu, 1989; Freedle & Kostin, 1988, 1991; Ironson & Subkoviak, 1979; Sinnott, 1980; Shepard, Camilli, & Williams, 1984; Schmitt & Bleistein, 1987; Schmitt & Dorans, 1990, in press). The purpose of the present research is to focus on the two examinee response style factors that seem to be salient for minority examinees-- differential speededness and differential omission-- and to evaluate their generalizability across minority focal groups and test content (Verbal and Mathematical). In addition, the interrelationships among differential omission, DIF, and item difficulty will also be studied.

Differential Item Functioning

Items that are harder for one group than for another group with the same level of ability or skill are defined as differentially more difficult or as functioning differentially between the two groups. Usually the majority group is referred to as the reference or base group and the minority group as the focal or study

group. Since DIF indices take overall differences in ability into account by matching the groups before comparing their item performance, DIF indices identify items that might have construct-irrelevant characteristics. Judgmental evaluation of items with DIF may identify some possible causes of DIF.

Examinee Response Style Factors

Examinee response style factors that may be related to differential item functioning deal with how different groups' examinees approach the test taking experience and how they deal with difficult items. These factors are differential speededness and differential omission.

Differential Speededness

When an examinee does not respond to an item and does not respond to any subsequent items in a timed section, all those items are referred to as not reached. In contrast, when an examinee does not respond to an item, but responds to subsequent items, that item is referred to as an omit. It is assumed that all examinees who do not respond to omitted items do so because the items are deemed too difficult. While some examinees may omit not-reached items because they are difficult, it is assumed that most, if not all, examinees who do not respond to a not-reached item, do not, in fact, reach the item. Not-reached rates on items at the end of a timed section or test measure the section or test speededness for the group that took the test.

Differential speededness refers to the existence of differential response rates between focal group members and matched reference group members to items appearing at the end of a section. Schmitt and Bleistein (1987) found evidence of this phenomenon for Black examinees, as compared to a matched group of White examinees, on analogy items. Schmitt and Dorans (1990) reported this effect for Hispanic examinees as well. Dorans, Schmitt, and Bleistein (1988) examined differential speededness results for Black, Hispanic, and Asian-American focal groups, compared to a White base or reference group. Results from that study indicated higher standardized differential not-reached rates on the last ten items of the two verbal sections of the two SAT-Verbal form studied for Black, Mexican-American, and Puerto Rican focal groups. No differential not-reached rates were observed for the Asian-American focal group.

Differential Omission

As stated before, when an examinee does not respond to an item, but responds to subsequent items, that item is referred to as an omit. Rivera and Schmitt (1988) assessed whether Puerto Rican and Mexican-American examinees were more or less likely to omit than matched White examinees. As with the samples used in the differential speededness studies, all examinees in the differential omit analysis spoke English as their best language. Standardization analyses were performed on omits using the version of the standardized difference in which all data including examinees who did not reach the item were used in the analysis.

The results of that study were somewhat surprising because they contradicted conventional wisdom which held that Hispanic examinees omit more than White examinees, a conventional wisdom based on comparisons of groups that were not matched on total test. After matching on total test score, Rivera and Schmitt (1988) found that Mexican-American examinees were less likely to omit than their matched White counterparts (in other words, they had more items with negative differential omission-- omitted differentially less). For Mexican-American examinees, more than twice as many items were identified as being omitted differentially less than as being omitted differentially more, and five items had more extreme negative differential omission, more negative than -.05. Interestingly, all five items contained true cognates. The results for Puerto Rican examinees were less conclusive since Puerto Rican examinees tended to omit slightly less than White examinees on Antonyms and Reading Comprehension items, but slightly more on Sentence Completion and Analogies, producing a cancellation effect of no differential omission on the test overall.

The Rivera and Schmitt (1988) study was the first of its nature and involved only one test form. In a subsequent study Schmitt and Bleistein (1987) found that the proportion of omits is related to item difficulty for White examinees. As part of a study investigating the relationship between item difficulty and DIF on the key, as measured by Mantel-Haenszel and standardization, Kulick and Hu (1989) recently corroborated the Rivera and Schmitt (1988) findings for Hispanic examinees (Mexican-American, Puerto Rican, and Latin American groups combined). Based on an analysis that included nine SAT test forms, they found that Hispanic examinees tend to omit less than matched Whites. Incidentally, they also found that DIF and difficulty are highly related for Hispanic, Black, and Asian-American groups: positive DIF tends to occur on hard items, while negative DIF tends to occur on easy items.

DIF, Item Difficulty, Differential Speededness, and Differential Omission

The relationships among DIF, item difficulty, differential speededness, and differential omission were studied for separate Hispanic groups (Mexican-American and Puerto Rican examinees) and by item type in two SAT-Verbal forms by Schmitt and Dorans (in press). They found that the most significant results occur with Analogy items. The most consistent finding, was, as for the other vocabulary item type, Antonyms, a strong negative relationship between DIF on the key and differential omission. In other words, items that were differentially harder (negative DIF) tended to be omitted differentially more (positive differential omission) and items that were differentially easier (positive DIF) tended to be omitted differentially less (negative differential omission) by Hispanic focal groups. In addition, a consistent and strong negative relationship between item difficulty and DIF on the key was also found to occur across both Hispanic groups on one form, while only the Puerto Rican focal group had this negative relationship on a second form. Conversely, both groups were found to exhibit a strong positive relationship between item difficulty and differential omission on this second form, but only the Mexican-American focal group had this positive relationship on the first form.

The most striking finding was the strong negative relationship between DIF on the key and differential omission that occurred across both forms for both groups on Antonyms and Analogies. Hispanics were found to be less likely to omit than Whites on items which have positive DIF, items which are relatively easier for them than for a group of matched Whites. The authors caution that:

> Before concluding that Hispanics should omit less in order to have more positive DIF, it is important to realize that this strong relationship may be somewhat artifactual for the following two reasons. First, there is a mathematical dependency that exists between selecting the key and omitting an item that may contribute to this negative correlation. Secondly, and perhaps more importantly, both the Mantel-Haenszel and the common version of the standardization index treat omits as wrong answers in their calculations, which may give spuriously high (in magnitude) DIF values on items in which differential omission occurs. In sum, this strong relationship between differential omission and DIF on the key may be due, in part, to a mathematical dependency and statistical artifact caused by erroneous treatment of omitted responses by the DIF indices in current use. (Schmitt & Dorans, in press, p.26)

Kulick and Hu (1989) also studied the interrelationship between DIF, item difficulty, and differential speededness and omission with data from nine other SAT forms. The main focus of the Kulick and Hu (1989) paper was to examine the relationship between item difficulty and DIF. To this purpose the authors tested

DIF prediction models which incorporated item difficulty and discrimination, item type, and differential not reached and omit indices. The authors concluded that item difficulty is related to DIF in that the more difficult items tend to be differentially easier for the minority focal group than for a comparable White reference group. This finding is consistent with the trend observed by the researchers referenced previously. An important finding by Kulick and Hu (1989) was that differential omission was even more highly related to DIF than item difficulty. Although differential speededness and differential omission have been documented by several research papers, its generalizability across different focal groups and tests has not been substantiated. Moreover, the interrelationship of these examinee response style factors to other item factors needs further study.

Purpose of Present Study

The purpose of this study is to evaluate differential item omit and speededness patterns obtained by different ethnic and gender groups on the Scholastic Aptitude Verbal and Mathematical Tests (SAT-Verbal and SAT-Math). Specifically this study will address the following issues: 1) The generalizability of the differential speededness and differentially lower omit patterns by Hispanic examinees to other ethnic and gender groups; 2) Whether using a non-speeded matching criterion affects DIF assessment for items that appear in the unspeeded sections of the test; and 3) The relationships among differential omission, item difficulty, and DIF.

Method

Instrument

Items on two Verbal and Mathematical forms of the Scholastic Aptitude Test (SAT) administered in 1984 and 1985 were used in the analyses for this study. The SAT-Verbal (SAT-V) test consists of 85 items administered in two separately-timed 30-minute sections. SAT-V items consist of four item types: Antonyms (ANT), Analogies (ANL), Reading Comprehension (RC), and Sentence Completion (SC). Within the SAT-V sections, items are ordered in the following manner:

Section 1: 15 ANT, 10 SC, 10 RC, & 10 ANL

Section 2: 10 ANT, 5 SC, 10 ANL, & 15 RC

The SAT-Mathematical (SAT-M) test score is composed of 35 five-option regular mathematics (RM) items

and 25 four-option Quantitative Comparison (QC) items. As with the SAT-V, the SAT-M is administered in two separately-timed 30-minute sections. Whereas Section 1 is comprised entirely of 25 RM items, Section 2 contains both mathematical item types and has the following item order:

Section 2: 7 RM, 25 QC, and 8 RM

Speededness

The speededness of a test is a measure of the degree to which test-takers were able to attempt all test items within the allotted time period. Several indices of speededness were evaluated for the 1984 and 1985 November SAT-V and SAT-M forms. The criterion used at ETS to indicate speededness in a test section is that all examinees are able to complete 75% of the section, and that 80% of the examinees are able to complete the entire section.

Over 99% of the examinees were able to complete 75% of Section 2 of the SAT-V test in 1984 and 1985. In contrast, only 97% of the examinees were able to complete 75% of Section 1 of the SAT-V test in either administration, which indicates that Section 1 is somewhat speeded. For the SAT-M test, nearly all examinees were able to complete 75% of Sections 1 and 2 in 1984, and Section 2 in 1985. Approximately 99% completed 75% of test items in Section 1 in 1985.

In order to assess whether 80% of the examinees completed the section, two indices are considered: the actual percentage completing the section, and the number of items reached by 80% of the examinees. The percentage of examinees completing the test may be low because the last item or several items in the test are speeded, or it may be low simply because the last item on the test is difficult. The number of items completed by 80% of the examinees is an index of the number of relatively unspeeded items in the test.

For the SAT-V test, 65.3% and 65.5% completed Section 1 & 2, respectively, in 1984. In 1985, the corresponding percentages were 76.7% and 71.8%. Eighty percent of the examinees were able to complete 41 of the 45 items in Section 1, and 37 of the 40 items in Section 2 of the SAT-V test administered in 1984. By comparison, 42 of the 45 items in Section 1, and 39 of the 40 items in Section 2 of the SAT-V test administered in 1985 were completed by 80% of the examinees. Therefore, it appears that both sections on the 1984 SAT-V test were more speeded than their counterparts on the 1985 test.

All items in Section 1 of the 1985 SAT-M test were completed by 80% of the examinees. All but one of the items were reached by 80% of the examinees in Section 2 in 1985, and in both sections of the SAT-M tests for 1984. Therefore, low completion rates for any section are probably due to the difficulty of the last item in the section. In 1984, only 61.4% of the examinees completed Section 1, while 72.8% were able to complete Section 2. In 1985, 80% completed Section 1, and 75.3% completed Section 2 of the SAT-M test.

Whereas both SAT-V forms were of comparable difficulty, the 1984 SAT-M form was more difficult than the 1985 form. Examinees taking the SAT were similar in ability, on average, for the two testing years.

Samples

Examinees taking a 1984 and a 1985 SAT form served as the populations for this study. From this pool, members of the following four ethnic groups were selected: Asian-American (n_{1984}= 11,417 & n_{1985}= 9,310), Black (n_{1984}= 22,250 & n_{1985}= 20,849), Hispanic (n_{1984}= 6,653-- comprised of approximately half Mexican-Americans and half Puerto Ricans-- & n_{1985}= 7,935-- comprised of approximately equal numbers of Latin Americans, Mexican-Americans, and Puerto Ricans), and White (n_{1984}= 280,520 & n_{1985}= 264,032). Members of both genders were also selected: males (n_{1984}= 149,373 & n_{1985}= 137,801) and females (n_{1984}= 181,400 & n_{1985}= 172,791). These groups represented the focal and reference groups used in the DIF analyses. The White group served as the reference group for all other ethnic focal groups, and males as the reference group for the gender comparison. All focal and reference groups used in this study contained only those examinees for whom English was their best language (1984) or their first language (1985).

Statistics

Two statistical procedures currently used at Educational Testing Service to assess DIF are the Mantel-Haenszel (MH) (Holland & Thayer, 1988) and the standardization (STD) methods (Dorans & Kulick, 1986). Both of these methods identify DIF after partitioning the reference and focal groups into subgroups with the same score on a relevant matching variable. The matching variable is usually the total score on a test closely related to the construct that the item is intended to measure. While there are some differences between the MH and STD methods, such as the scale in which the item performance of the reference and focal groups are compared and the way that the individual differences between the subgroups are averaged, the DIF

estimates computed by these methods are highly correlated (upper .90's) because they tend to yield the same rank order of items with respect to DIF (Wright, 1987; Holland & Thayer, 1988; Dorans, 1989).

Standardization procedure. In the traditional standardization analysis, an item is said to exhibit DIF when the probability of correctly answering the item is lower or higher for examinees from one group than for equally able examinees from another group. The focus of DIF analyses is on differences in performance between groups that are matched with respect to the ability, knowledge, or skill of interest. The basic elements of a standardization analysis of the keyed response are proportions correct at each level of a matching variable, such as total score, in a base or reference group and a focal or study group. Plots of these conditional proportions correct against score level in the focal and reference groups provide a visual indication of the extent of DIF that an item exhibits. A plot of differences in conditional proportions correct between the focal and reference group portrays the degree of DIF more directly. In addition to these plots, standardization provides numerical indices for quantifying DIF.

The prime numerical DIF index that standardization computes is the standardized p-difference, which is defined as

(1) $$DSTD = \Sigma\{W_s[P_{fs} - P_{rs}]\} / \Sigma\{W_s\},$$

where $[W_s / \Sigma\{W_s\}]$ is the weighing factor at score level s on the SAT used to weight differences in the proportions correct between the focal group (P_{fs}) and the reference group (P_{rs}), and Σ is the summation operator which sums these weighted differences across scores levels to arrive at DSTD, an index that can range from -1 to +1 or -100% to 100%. Negative values of DSTD indicate that the item disadvantages the focal group, while positive DSTD values indicate that the item favors the focal group. DSTD values between -.05 (-5%) and +.05 (+5%) are considered negligible. DSTD values outside the {-.10, +.10} or the {-10%, +10%) range are considered sizeable. For operational purposes, a $|DSTD| \geq .10$ is a recommended cutoff; for research purposes, a cutoff of $|DSTD| \geq .05$ should be used.

The weights, $[W_s / \Sigma\{W_s\}]$, which are applied to both P_{fs} and P_{rs}, are the essence of the standardization approach. Contrast this weighing constancy with what occurs in the computation of impact,

(2) $$IMPACT = P_f - P_b = \Sigma\{N_{fs}P_{fs}\}/\Sigma\{N_{fs}\} - \Sigma\{N_{rs}P_{rs}\}/\Sigma\{N_{rs}\},$$

where N_{fs} and N_{rs} are the frequencies of score level s in the focal and reference groups. Thus, impact provides differences in performance between groups which have not been matched and <u>are probably not</u> comparable. In contrast, the standardization method focuses on comparing groups that <u>are</u> comparable. In addition, the particular set of weights employed for standardization depends upon the purposes of the investigation. In practice, $W_s = N_{fs}$ has been used because it gives the greatest weight to differences in P_{fs} and P_{rs} at those score levels most often attained by the focal group under study. Use of N_{fs} means that DSTD equals the difference between P_f, the observed performance of the focal group on the item, and P_f, the imputed performance of selected reference group members who are matched in ability to the focal group members.

In research on the SAT, two versions of DSTD have been computed. In the original version, all examinees including those who do not reach the item are included in the denominator of P_{fs} and P_{rs}, yielding $DSTD_1$. In the more recent version, an effort was made to adjust for speededness by excluding the not reached examinees from the calculation of DSTD, yielding $DSTD_2$. Schmitt and Bleistein (1987) were the first to employ this correction, which has become the standard in operational DIF work on the SAT. Dorans, Schmitt, and Curley (1988) demonstrated that this correction partially adjusts for the speededness effect. Hence, $DSTD_2$ will be used in this report for all options except not reached where only $DSTD_1$ makes sense.

The generalization of the standardization methodology to all response options including omitting and not reaching the item is straightforward. It is as simple as replacing the keyed response with the option of interest in all calculations. For example, a standardized response rate analysis on omits would entail computing the proportions omitting (as opposed to the proportions correct) in both the focal and reference groups,

(3) $$P_{fs}(O) = O_{fs}/N_{fs};\ P_{rs}(O) = O_{rs}/N_{rs},$$

where O_{fs} and O_{rs} are the number of people in the focal and reference groups, respectively, at score level s who omit. The next step is to compute differences between these proportions,

(4) $$D_s(O) = P_{fs}(O) - P_{rs}(O).$$

Then these individual score level differences are summarized across score levels by applying some standardized

weighing function to these differences to obtain DSTD(O),

(5) $$DSTD(O) = \Sigma\{W_s[P_{fs}(O) - P_{rs}(O)]\} / \Sigma\{W_s\},$$

the standardized difference in omit rates. In a similar fashion one can compute standardized differences in response rates for options A, B, C, D, and E, and for non-reaches as well.

For non-reaches, application of the standardization methodology to counts of examinees at each score who did not reach the item culminates in a standardized not-reached difference,

(6) $$DSTD(NR) = \Sigma\{W_s[P_{fs}(NR) - P_{rs}(NR)]\} / \Sigma\{W_s\}.$$

For items at the end of a separately-timed section of a test such as the SAT, these standardized differences provide measurement of the differential speededness of a test. Differential speededness refers to the existence of differential response rates between focal group members and matched reference group members to items appearing at the end of a section.

Mantel-Haenszel method. The Mantel-Haenszel (Mantel & Haenszel, 1959) procedure, adapted by Holland and Thayer (1988) for DIF analysis computes ratios of the conditional odds of successful reference group performance over the conditional odds of successful focal group performance at each score level, and then averages these ratios across score levels. In the calculation of the average ratio, statistically optimal weights are used for each ratio. The Mantel-Haenszel statistic is transformed to the "delta" scale used in the ETS test development process. The resulting statistic is used for the estimation of a differential item functioning index for the SAT.

Formula score DIF. Both the standardized p-difference ($DSTD_1$ or $DSTD_2$) and the Mantel-Haenszel score items as right (1) or wrong (0). For formula-score tests, such as the SAT, a different item scoring procedure is more appropriate from a logical point of view, namely an item formula score version of standardization. Instead of scoring each item as correct or incorrect at each score level, one obtains the expected item formula score at each score level,

(7) $$E_{fs} = (R_{fs} - W_{fs}/(K-1))/N_{fs},$$

where R_{fs}, W_{fs}, and N_{fs} are the number of examinees getting the item right, the number getting it wrong, and

the total number of examinees in the focal group at score level s. K is the number of response options. For the reference group, one computes,

(8) $$E_n = (R_n - W_n/(K-1))/N_n .$$

Then E_{fs} and E_n can be used in place of P_{fs} and P_n in Equation 1 to obtain DSTD-FS. For a more complete discussion of standardized formula-score DIF, see Dorans and Holland (1989).

Procedure

The standardization method was used to compute differential omission rates, differential speededness, and differential item functioning for items on the SAT-Verbal and SAT-Mathematical test forms. Differential item functioning statistics were also calculated using the Mantel-Haenszel method. The analysis of DIF involved a two-step process to determine and compare two matching criteria. During the first step the matching criterion for the DIF analyses was the total-test raw score on either the SAT-V or the SAT-M form, henceforth referred to as the speeded matching criterion. On the basis of the initial analysis, the percentage of examinees from the reference and focal groups who did not reach the items were computed. An unspeeded criterion was determined for each of the two 1984 and 1985 SAT-Verbal or SAT-Math sections by dropping those items with a not-reached percentage greater than 5%, for most groups, from the total 85-item Verbal or 60-item Mathematical full-length criteria. Accordingly, the unspeeded matching criteria are as follows:

1984 SAT-V: Section 1 -- (45 items - 10 ANL) = 35
Section 2 -- (40 items - 5 RC) = 35
Unspeeded 1984 SAT-V Total = 70

1985 SAT-V: Section 1 -- (45 items - 3 RC & 10 ANL) = 32
Section 2 -- (40 items - 7 RC) = 33
Unspeeded 1985 SAT-V Total = 65

1984 SAT-M: Section 1 -- (25 items - 5 RM) = 20
Section 2 -- (35 items - 5 RM) = 30
Unspeeded 1984 SAT-M Total = 50

1985 SAT-M: Section 1 -- (25 items - 5 RM) = 20
Section 2 -- (35 items - 5 RM) = 30
Unspeeded 1985 SAT-M Total = 50

Using the unspeeded criteria a second set of analyses were computed. In this second analysis, new DIF statistics were obtained for items in all focal group/reference group comparisons. Values from the initial analysis (speeded criterion) were compared to the values obtained in the second analysis (unspeeded criterion).

Results and Discussion

Analyses will be summarized using descriptive statistics and graphical display of the omit and speededness patterns of women/men, Blacks/Whites, Asian-Americans/Whites, and Hispanic/Whites. Two administrations of the SAT-Verbal will be used to evaluate the replicability of results. All analyses are summarized by Verbal and Mathematical sections.

Differential Speededness

Results of the differential speededness effect found for the women/men, Black/White, Asian-American/White, and Hispanic/White comparisons are displayed in Figure 1 for the two SAT-Verbal Forms (two sections each) and on Figure 2 for the two SAT-Math Forms (two sections each).

Insert Figures 1 and 2 about here

Examination of the not-reached standardized differences for the last ten items of each SAT-Verbal section show that Black and Hispanic groups have a higher not-reached proportion than the matched White reference group. More extreme differential speededness are observed for both of these focal groups for the SAT-V Section 1 (the more speeded 45-item section) than for the SAT-V Section 2 (the 40 item section). Nevertheless, a consistent pattern of differential speededness is observed across both sections. Black examinees have the highest proportion of differential not-reached, followed by the Hispanic focal group. Not much differential speededness is observed for, either, the Asian-American focal group or for the female focal group. In fact, a slight positive differential speededness pattern seems to appear for these two focal groups; Asian-Americans and females reach slightly more items at the end of the Verbal sections than their respective matched White or male groups. This pattern is slightly more prominent for the female focal group on the SAT-V Section 2. Results for the SAT-Math sections indicate a very similar pattern for the Black and Hispanic focal groups. Nevertheless, the slight positive differential speededness (negative not-reached)

observed in the Verbal sections for the Asian-American and female focal groups is not evident for the Math sections. A slight negative differential speededness is observed for the Asian-American focal group, particularly at the very end of the Math 2 sections. The findings across the ethnic groups are consistent with those reported by Schmitt and Dorans (1990). The findings for the female/male comparison have not been reported before.

Speeded versus Unspeeded Criterion

In order to determine if using an unspeeded criterion would change the results obtained with a speeded criterion, all DIF analyses were done using both criteria. Results of these two computations are presented in Table 1 and 2 for the SAT-Verbal and SAT-Math sections respectively.

Insert Tables 1 and 2 about here

Tables 1 and 2 present descriptive statistics (means, standard deviations, and intercorrelations for the three DIF statistics computed: the standardization DSTD in the p metric, the formula score DIF, DSTD-FS in the formula-score metric, and the Mantel-Haenszel, MH D DIF in the delta metric, and for the standardization differential omission (DIF omit), across the two matching criteria. Only data for unspeeded items were included in the calculations. Comparison of the mean and standard deviations of the statistics under the two matching criteria conditions indicate that the DIF statistics and DIF omit are unaffected by the speededness of the matching criterion. The correlations between the indices under the speeded and unspeeded conditions indicate that the relative ordering of items for all three DIF indices and for the differential omission are parallel since all correlations are positive and almost perfect (1.00). Based on these results, it appears that matching groups on the basis of a speeded or unspeeded criterion does not affect the estimation of DIF or differential omission indices for the unspeeded items. One basic consideration of these results is that none of the tests in this study had extreme speededness and, consequently, the unspeeded criteria were nearly full length tests. It is possible that if the speeded criterion is extremely speeded, these results would be different. Because the indices obtained under the two matching criteria were basically the same, the subsequent analyses were restricted to the estimates of the unspeeded items only obtained after matching on the full total test (speeded) criterion.

Relationships Among Differential Omission, Item Difficulty, and DIF

Correlations between the three DIF indices, differential omission, and item difficulty are presented in Tables 3 and 4. In addition, these tables also present means and standard deviations for these statistics.

Insert Tables 3 and 4 about here

Results presented in these tables indicate that the DSTD-P and DSTD-FS estimates are basically the same and that their correlations with the MH D DIF statistic is consistently high and positive (.87 to .99) across tests (Verbal or Mathematical), forms (1984 or 1985), and focal comparison groups (females, Blacks, Asian-Americans, and Hispanics).

The relationship between DIF indices and item difficulty (P%) is mostly negative (more negative DIF for easier items) but not as strong as found by Kulick and Hu (1989). Interestingly, a slight positive relationship between DIF and P% was found for the female/male comparisons for the Math sections of the 1984 SAT. Scatterplots of DSTD-P and item difficulty are presented in Figures 3 and 4 for the Verbal and Mathematical sections respectively. No clear or consistent pattern between DSTD-P and item difficulty is apparent in these plots for the reduced unspeeded items of the Verbal or Mathematical SAT sections. In addition, only a handful of items had DSTD-P values that exceeded .10 either in a positive or negative direction for any of the groups studied.

Insert Figures 3 and 4 about here

The relationships between differential omit (DIF Omit) and P% are presented in Tables 3 and 4 and displayed on Figures 5 and 6 for the Verbal and Mathematical tests, respectively. Positive relationships between DIF Omit and P% are observed for both forms of the Verbal tests for the female/male, Black/White, and Hispanic/White comparisons. Females omit less than matched males on harder items, while Blacks and Hispanics omit less than matched Whites on harder items. Negative correlations were obtained for the Asian-American/White comparisons across both Verbal forms and for the female/male comparisons across both Math forms. Asian-Americans omit more than matched Whites on harder verbal items across both forms, while females omit more than matched males on harder math items across both forms.

Insert Figures 5 and 6 about here

These results may indicate a differential test wiseness by the focal groups with the Asian-American group demonstrating a higher tendency to omit the more difficult Verbal items than that demonstrated by the White matched comparison group. All the other focal groups appeared to have a lower omission of the harder Verbal items than did the matched comparison groups.. These results have a parallel with the differential speededness observed for the ethnic groups, where the Black and Hispanic comparison groups do not reach the last items of the sections to a higher degree than matched White groups. It is possible that a tendency to omit differentially less (and with less discrimination for the difficulty of the item) slows their progress and prevents them from completing the same number of items as do comparable White examinees. Results for the Mathematical sections are not as consistent except for the female/male comparison. As stated previously, the relationship found for the Mathematical sections for the female/male comparison indicated that females may omit the more difficult items at a higher rate than do comparable male examinees.

Differential Omission

Differential omission means and standard deviations presented in Tables 3 and 4 indicate a small but consistent positive DIF Omit for Asian-Americans on the Verbal portions of the two SAT tests studied. When the DIF Omit values are plotted against item difficulty (Figures 5 and 6), it is interesting to note that, particularly for the Verbal sections, the larger differential omissions occur for the harder items. Females and Blacks (and to a slightly lesser degree, Hispanics) omit the harder items differentially less than their respective matched reference groups, while Asian-Americans omit the same harder items differentially more than the White reference group. No similar pattern is apparent for the Mathematical sections.

Figures 7 and 8 present the relationship between DIF Omit and DSTD-P for the Verbal and Mathematical sections, respectively. As noted in the correlations in Tables 3 and 4 and in the scatter plots of Figures 7 and 8, the relationship between these two variables is consistently high and negative.

Insert Figures 7 and 8 about here

This negative association had previously been reported by Rivera and Schmitt (1988), Kulick and Hu (1989), and Schmitt and Dorans (in press). As pointed out by these authors, the relationship between DIF Omit and DIF is artifactual because of the ipsative nature of the calculations. A cautionary explanation about the artifactual relationship between these two variables was provided by Schmitt and Dorans (in press, p.26) and was quoted on page 5 of the Introduction for this paper. In addition, Kulick and Hu (1989) expand on the ipsative nature of the relationship:

> That is, MHD (MH D DIF) and DIFPOM (DIF Omit) are *dependent*. Observe that all response percentages must sum to 100 percent. If omitting and not reaching an item are merely considered two additional response alternatives, then the sum across all response percentages *including omits and the correct response is 100 percent*. This is true whether the response percentages are standardized (as they are here) or not. (Kulick & Hu, 1989, p.8)

They further summarize:

> Although DIF and DIFPOM are dependent measures because of their ipsative relationships, there is no reason why positive DIF must be counterbalanced with negative differential omitting, rather than negative DIF on the item's distractors. To what extent DIF is a consequence of differential omission and to what extent differential omission is a manifestation of DIF is problematic. Nonetheless, the presence of differential omission on a test has the *potential* to influence DIF indices and therefore should be an important concern. (Kulick & Hu, 1989, p. 9)

Conclusions

The present study examined three issues pertaining to DIF and differential non-response. First, it examined whether differential speededness and differential omission extended across ethnic groups, SAT tests, and test forms. Second, it examined the effects of using an unspeeded matching criterion on the DIF assessment of items that are not affected by speed. Third, it examined the complex interrelationships among DIF, differential omission, and item difficulty. In addition, this study examined, as a methodological side issue, whether the results obtained depended on whether one used Mantel-Haenszel, standardized p-differences or standardized formula-score differences. As it turned out, it did not matter which index was used because they were very highly correlated.

With respect to the first of the three substantive issues, clear replicable evidence was found for differential speededness for Blacks and Hispanics across both Math tests and both Verbal tests. Blacks and Hispanics do not reach the items at the end of sections as often as matched Whites do. In contrast, there

appears to be little consistent evidence of differential speededness for females compared to matched males or for Asian-Americans compared to matched Whites.

The results for differential omission are more complex. As expected, differential omission depends on item difficulty. On easy items, there was little omission by any of the groups studied. Hence, there was little opportunity for differential omission. On hard items, differential omission occurred. Here, Black and females, and to a lesser degree, Hispanics, omit differentially less than their reference groups, while Asian-Americans omit differentially more than their reference groups.

Assessing the effects of a speeded criterion on DIF assessment is very straightforward for the items that are part of the unspeeded criterion. This can be accomplished by dropping the speeded items from the criterion and redoing the DIF analysis without them, as was done in this study. The results of this study are convincing in demonstrating that the speeded and unspeeded criteria are exchangeable for purposes of analyzing the unspeeded items. This result is probably due to the long length of the matching variable and the relative unspeeded nature of the tests we studied.

Assessing the effects of a speeded versus unspeeded criterion on the DIF assessment of the speeded items is not straightforward, and we made no attempt to do it. The reason it is not straightforward is that exclusion of an item from the criterion in a DIF analysis will lead to the appearance of DIF on that item when in fact there is no DIF (Holland & Thayer, 1988). Hence, we need to include each item, one at a time, in the criterion. (If we included all of them at once, we would have the speeded criterion.) Unfortunately, there is no known way to correctly include an item into a formula-scored criterion. For right-scored tests, inclusion is straightforward, but assessment of speededness is not. Since the SAT is a formula-scored test, we chose to study that which we could, the unspeeded items, and concluded that the speededness of the criterion is unimportant for the tests in this study.

The DIF/difficulty relationship observed by Kulick and Hu (1989) was observed, but not as strongly in this study. This relationship may have been attenuated because the speeded items were not included in the analyses. These items tend to be difficult, and hence exclusion of them may attenuate the relationship difficulty can have with other variables.

This expected attenuation should not occur for the difficulty and differential omission relationship, however, because the speeded difficult items that have little omission are excluded from the analysis. And it does not occur. The positive difficulty and differential omission relationship is fairly strong across all groups on the Verbal test except for Asian-Americans for whom the relationship is negative. The only consistent correlations between difficulty and differential omission on the Mathematical test occurred for females and was negative.

Finally, the artifactual relationship between DIF and differential omission that Kulick and Hu (1989) and Schmitt and Dorans (in press) demonstrated is replicated here. Whenever there is DIF, there must be "anti-DIF" since the sum of the standardized differences across distractors, the key, and non-responses has to be zero. When an examinee does not know the answer to a difficult question, the examinee, if well-prepared, will tend to omit. If the examinee is not well-prepared, he or she may tend to respond even when he or she does not know the answer to a hard test question. Some DIF may simply reflect the differential test preparedness of different subgroups of examinees. Given the different levels of test preparedness that may exist across subgroups of test-takers, it is not surprising that much "anti-DIF" takes on the form of differential omission. Research in which levels of test preparedness are varied is needed to help resolve whether DIF is due to differential omission or whether differential omission results from DIF in the item.

References

Alderman, D. L., & Holland, P. W. (1981). *Item performance across native language groups on the test of English as a foreign language* (RR-81-16). Princeton, NJ: Educational Testing Service.

Bleistein, C. A., Schmitt, A. S., & Curley, W. E. (1990, April). *Factors affecting the performance of Black examinees on SAT analogy items*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, Boston, MA.

Bleistein, C. A., & Wright, D. J. (1987). Assessing unexpected differential item difficulty for Asian-American examinees on the Scholastic Aptitude Test. In A. P. Schmitt & N. J. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (RM-87-1). Princeton, NJ: Educational Testing Service.

Breland, H. M., Stocking, M., Pinchak, B. M., & Abrams, N. (1976). *The cross-cultural stability of mental test items: An investigation of response patterns for ten sociocultural groups* (RR-74-2). Princeton, NJ: Educational Testing Service.

Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing, 2*, 155-163.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education, 2*, 217-233.

Dorans, N. J., & Holland, P. W. (1989, October). *DIF detection and description: Mantel-Haenszel and standardization*. Paper presented at the ETS Conference, Differential Item Functioning: Theory and Practice, Princenton, NJ.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368.

Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1988). *The standardization approach to assessing differential speededness* (RR-88-31). Princeton, NJ: Educational Testing Service.

Dorans, N. J., Schmitt, A. P., & Curley, W. E. (1988, March). *Differential speededness: Some items have DIF because of where they are, not what they are*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Freedle, R., & Kostin, I. (1991). *Semantic and structural factors affecting the performance of matched Black and White examinees on analogy items from the Scholastic Aptitude Test* (RR-91-28). Princeton, NJ: Educational Testing Service.

Freedle, R., & Kostin, I. (1988). *Relationship between item characteristics and an index of differential item functioning (DIF) for the four GRE verbal item types* (RR-88-29). Princeton, NJ: Educational Testing Service.

Green, B.F., Crone, C. R., & Folk, V. G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement, 26*, 147-160.

Holland, P. W., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Erlbaum.

Ironson, G. H., & Subkoviak, M. J. (1979). A comparison of several methods of assessing item bias. *Journal of Educational Measurement, 16*, 209-225.

Kulick, E., & Dorans, N. J. (1983). *Assessing unexpected differential item performance of Oriental candidates on SAT Form CSA6 and TSWE Form E33* (SR-83-106). Princeton, NJ: Educational Testing Service.

Kulick, E., & Hu, P. G., (1989). *Examining the relationship between differential item functioning and item difficulty* (CB-89-5). New York, NY: College Entrance Examination Board.

Mantel, N., & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Rivera, C., & Schmitt, A. P. (1988). *A comparison of Hispanic and White students' omit patterns on the Scholastic Aptitude Test* (RR-88-44). Princeton, NJ: Educational Testing Service.

Scheuneman, J. D. (1987). An experimental exploratory study of causes of bias in test items. *Journal of Educational Measurement, 24*, 97-118.

Scheuneman, J. D., & Bleistein, C. A. (1989). A consumer's guide to statistics for identifying differential item functioning. *Applied Measurement in Education, 2*, 255-275.

Schmitt, A. P. (1985). *Assessing unexpected differential item performance of Hispanic candidates on SAT Form 3FSA08 and TSWE Form E47* (SR-85-169). Princeton, NJ: Educational Testing Service.

Schmitt, A. P. (1988). Language and cultural characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. *Journal of Educational Measurement, 25*, 1-13.

Schmitt, A. P., & Bleistein, C. A. (1987). *Factors affecting differential item functioning for Black examinees on Scholastic Aptitude Test analogy items* (RR-87-23). Princeton, NJ: Educational Testing Service.

Schmitt, A. P., Curley, W. E., Bleistein, C. A., & Dorans, N. J. (1988, March). *Experimental evaluation of language and interest factors related to differential item functioning for Hispanic examinees on the SAT-Verbal.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Schmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement, 27*, 67-81.

Schmitt, A. P., & Dorans, N. J. (in press). Factors related to differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. In G. Keller, et al. (Eds.), *Assessment and Access: Hispanics in Higher Education*, SUNY Press.

Shepard, L. A., Camilli, G., & Williams, D. M. (1984). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics, 9*, 93-128.

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement, 22*, 77-105.

Sinnott, L. T. (1980). *Differences in item performance across groups* (RR-81-16). Princeton, NJ: Educational Testing Service.

Wright, D. J. (1987). An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. In A. P. Schmitt & N. J. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (RM-87-1). Princeton, NJ: Educational Testing Service.

Table 1A

Descriptive Statistics for DIF and Differential Omission When Using Speeded and Unspeeded Matching Criteria for Verbal Sections of the 1984 SAT[1]

	Section 1[2]								Section 2[2]							
	Speeded		Unspeeded		Correlations[3]				Speeded		Unspeeded		Correlations[3]			
	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit
Female/Male																
DSTD-P	.00	.04	-.00	.04	1.00	1.00	.97	-.70	.01	.05	.00	.05	1.00	1.00	.98	-.72
DSTD-FS	.01	.05	-.00	.05	1.00	1.00	.97	-.70	.01	.06	.00	.06	1.00	1.00	.98	-.72
MH D DIF	.06	.55	-.01	.55	.97	.97	1.00	-.68	.15	.82	.07	.82	.98	.98	1.00	-.65
DIF Omit	-.01	.02	-.01	.02	-.69	-.69	-.67	.99	-.01	.02	-.01	.02	-.71	-.71	-.64	.99
Black/White																
DSTD-P	.01	.03	.01	.03	1.00	1.00	.95	-.55	-.00	.04	-.01	.04	1.00	1.00	.98	-.46
DSTD-FS	.01	.04	.01	.04	1.00	1.00	.95	-.55	-.00	.05	-.01	.05	1.00	1.00	.98	-.46
MH D DIF	.16	.45	.13	.47	.96	.96	1.00	-.58	-.06	.59	-.10	.60	.99	.99	1.00	-.55
DIF Omit	-.00	.01	-.00	.01	-.53	-.53	-.58	.98	-.01	.02	-.01	.02	-.51	-.51	-.59	.99
Asian American/White																
DSTD-P	-.00	.04	-.00	.04	1.00	1.00	.94	-.52	-.01	.04	-.01	.04	1.00	1.00	.92	-.39
DSTD-FS	-.00	.05	-.00	.05	1.00	1.00	.94	-.52	-.01	.04	-.01	.04	1.00	1.00	.92	-.39
MH D DIF	-.10	.64	-.09	.65	.94	.94	1.00	-.38	-.16	.60	-.15	.60	.92	.92	1.00	-.24
DIF Omit	.02	.02	.02	.02	-.53	-.53	-.39	.99	.02	.02	.01	.02	-.38	-.38	-.24	.99
Hispanic/White																
DSTD-P	.01	.03	.01	.03	1.00	1.00	.96	-.62	.00	.04	-.00	.04	1.00	1.00	.97	-.58
DSTD-FS	.01	.04	.01	.04	1.00	1.00	.96	-.62	.00	.04	-.00	.04	1.00	1.00	.97	-.58
MH D DIF	.07	.48	.04	.50	.97	.97	1.00	-.64	-.01	.55	-.05	.57	.97	.97	1.00	-.63
DIF Omit	-.01	.01	-.01	.01	-.61	-.61	-.63	.98	-.01	.02	-.01	.02	-.59	-.59	-.64	.99

[1]The speeded criterion is based on all 85 items across both SAT Verbal sections. The unspeeded criterion is based on 70 items: 35 items in section 1 and 35 items in section 2.

[2]Only data for unspeeded items were included in these calculations (section 1: 35 items and section 2: 35 items).

[3]In the correlation matrix, DIF and DIF Omit variables based on the speeded criterion are presented in the rows, and those based on the unspeeded criterion are in the columns. All correlations are across-criteria correlations.

Table 1B

Descriptive Statistics for DIF and Differential Omission When Using Speeded and Unspeeded Matching Criteria for Verbal Sections of the 1985 SAT[1]

	Section 1[2]								Section 2[2]							
	Speeded		Unspeeded		Correlations[3]				Speeded		Unspeeded		Correlations[3]			
	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit
Female/Male																
DSTD-P	.01	.04	.01	.04	1.00	1.00	.95	-.73	-.00	.04	-.00	.04	1.00	1.00	.88	-.63
DSTD-FS	.01	.05	.01	.05	1.00	1.00	.95	-.73	-.00	.04	-.00	.04	1.00	1.00	.88	-.63
MH D DIF	.05	.59	.04	.59	.94	.94	1.00	-.75	-.07	.74	-.08	.73	.87	.87	1.00	-.50
DIF Omit	-.02	.03	-.02	.03	-.72	-.72	-.74	.99	-.01	.02	-.01	.02	-.65	-.65	-.52	.99
Black/White																
DSTD-P	.01	.03	.01	.03	1.00	1.00	.95	-.43	-.00	.05	-.00	.05	1.00	1.00	.97	-.70
DSTD-FS	.01	.04	.01	.04	1.00	1.00	.95	-.43	-.00	.06	-.00	.06	1.00	1.00	.97	-.70
MH D DIF	.13	.46	.08	.48	.95	.95	1.00	-.44	.00	.71	-.05	.73	.98	.98	1.00	-.69
DIF Omit	-.00	.01	-.00	.01	-.47	-.47	-.47	.97	-.00	.02	-.00	.02	-.72	-.72	-.70	.99
Asian American/White																
DSTD-P	-.00	.03	.00	.03	1.00	1.00	.95	-.46	-.00	.03	-.01	.03	1.00	1.00	.96	-.55
DSTD-FS	-.00	.04	-.00	.04	1.00	1.00	.95	-.46	-.01	.04	-.01	.04	1.00	1.00	.96	-.55
MH D DIF	-.04	.47	-.05	.48	.95	.95	1.00	-.39	-.04	.51	-.05	.51	.96	.96	1.00	-.50
DIF Omit	.02	.02	.02	.02	-.44	-.44	-.35	.99	.02	.02	.01	.02	-.53	-.53	-.49	.98
Hispanic/White																
DSTD-P	.01	.02	.01	.02	1.00	1.00	.90	-.55	-.01	.03	-.01	.03	1.00	1.00	.97	-.74
DSTD-FS	.01	.03	.01	.03	1.00	1.00	.90	-.55	-.01	.04	-.01	.04	1.00	1.00	.97	-.74
MH D DIF	.21	.39	.17	.40	.89	.89	1.00	-.48	-.05	.49	-.09	.51	.98	.98	1.00	-.70
DIF Omit	-.00	.01	-.00	.01	-.58	-.58	-.52	.96	-.00	.01	-.00	.01	-.77	-.77	-.71	.98

[1]The speeded criterion is based on all 65 items across both SAT Verbal sections. The unspeeded criterion is based on 65 items: 32 items in section 1 and 33 items in section 2.

[2]Only data for unspeeded items were included in these calculations (section 1: 32 items and section 2: 33 items).

[3]In the correlation matrix, DIF and DIF Omit variables based on the speeded criterion are presented in the rows, and those based on the unspeeded criterion are in the columns. All correlations are across-criteria correlations.

23

Table 2A

Descriptive Statistics for DIF and Differential Omission When Using Speeded and Unspeeded Matching Criteria for Math Sections of the 1984 SAT[1]

	Section 1[2]								Section 2[2]							
	Speeded		Unspeeded		Correlations[3]				Speeded		Unspeeded		Correlations[3]			
	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit
Female/Male																
DSTD-P	.00	.04	-.00	.05	1.00	1.00	.98	-.86	.00	.03	.00	.03	1.00	1.00	.98	-.52
DSTD-FS	.00	.05	-.00	.05	1.00	1.00	.98	-.86	.00	.03	.00	.03	1.00	1.00	.99	-.52
MH D DIF	.03	.62	-.04	.63	.98	.98	1.00	-.90	.08	.37	.02	.38	.98	.98	1.00	-.51
DIF Omit	.01	.02	.01	.02	-.84	-.84	-.89	.99	.00	.01	.01	.01	-.58	-.58	-.59	.97
Black/White																
DSTD-P	-.01	.04	-.01	.04	1.00	1.00	.98	-.51	.00	.03	.00	.03	1.00	1.00	.98	-.47
DSTD-FS	-.01	.05	-.01	.05	1.00	1.00	.98	-.51	.00	.04	.00	.04	1.00	1.00	.98	-.46
MH D DIF	-.14	.50	-.11	.51	.98	.98	1.00	-.55	.03	.38	.06	.38	.98	.98	1.00	-.44
DIF Omit	-.00	.01	-.00	.01	-.56	-.57	-.60	.98	.00	.01	.00	.01	-.58	-.57	-.56	.96
Asian American/White																
DSTD-P	-.01	.05	-.01	.05	1.00	1.00	.96	-.76	.01	.03	.01	.03	1.00	1.00	.95	-.68
DSTD-FS	-.01	.05	-.01	.05	1.00	1.00	.96	-.76	.01	.04	.01	.04	1.00	1.00	.95	-.67
MH D DIF	-.17	.80	-.16	.80	.96	.96	1.00	-.70	.12	.51	.14	.52	.96	.96	1.00	-.65
DIF Omit	.01	.02	.00	.02	-.78	-.78	-.71	.99	-.00	.01	-.00	.01	-.69	-.68	-.66	.98
Hispanic/White																
DSTD-P	-.01	.02	-.00	.02	1.00	1.00	.99	-.63	.00	.02	.00	.02	1.00	1.00	.98	-.37
DSTD-FS	-.01	.03	-.01	.03	1.00	1.00	.99	-.63	.00	.02	.01	.02	1.00	1.00	.99	-.37
MH D DIF	-.09	.38	-.06	.38	.99	.99	1.00	-.61	.04	.27	.06	.27	.98	.98	1.00	-.41
DIF Omit	-.00	.01	-.01	.01	-.71	-.71	-.69	.97	-.00	.01	-.01	.01	-.34	-.34	-.37	.97

[1]The speeded criterion is based on all 60 items across both SAT Math sections. The unspeeded criterion is based on 50 items: 20 items in section 1 and 30 items in section 2.

[2]Only data for unspeeded items were included in these calculations (section 1: 20 items and section 2: 30 items).

[3]In the correlation matrix, DIF and DIF Omit variables based on the speeded criterion are presented in the rows, and those based on the unspeeded criterion are in the columns. All correlations are across-criteria correlations.

Table 2B

Descriptive Statistics for DIF and Differential Omission When Using Speeded and Unspeeded Matching Criteria for Math Sections of the 1985 SAT[1]

	Section 1[2]								Section 2[2]							
	Speeded		Unspeeded		Correlations[3]				Speeded		Unspeeded		Correlations[3]			
	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit	Mean	SD	Mean	SD	DSTD-P	DSTD-FS	MH D DIF	DIF Omit
Female/Male																
DSTD-P	-.00	.04	-.00	.04	1.00	1.00	.97	-.59	-.00	.04	-.00	.04	1.00	1.00	.98	-.77
DSTD-FS	-.00	.04	-.00	.04	1.00	1.00	.97	-.59	-.00	.05	-.00	.05	1.00	1.00	.98	-.76
MH D DIF	-.02	.57	-.03	.57	.97	.97	1.00	-.55	-.01	.58	-.01	.58	.98	.98	1.00	-.73
DIF Omit	.01	.02	.01	.02	-.63	-.63	-.60	.98	.01	.02	.01	.02	-.77	-.77	-.75	.99
Black/White																
DSTD-P	-.01	.04	-.01	.04	1.00	1.00	.99	-.70	.00	.03	.00	.03	1.00	1.00	.99	-.43
DSTD-FS	-.01	.05	-.01	.05	1.00	1.00	.99	-.70	.00	.04	.01	.04	1.00	1.00	.99	-.43
MH D DIF	-.14	.54	-.12	.54	.98	.98	1.00	-.68	.05	.44	.06	.44	.99	.99	1.00	-.47
DIF Omit	.00	.02	.00	.02	-.73	-.73	-.71	.99	.00	.01	-.00	.01	-.46	-.45	-.50	.97
Asian American/White																
DSTD-P	.00	.03	-.00	.03	1.00	1.00	.98	-.77	-.00	.03	-.00	.03	1.00	1.00	.96	-.71
DSTD-FS	.00	.03	-.00	.03	1.00	1.00	.98	-.77	-.00	.03	-.00	.03	1.00	1.00	.96	-.71
MH D DIF	.02	.45	.01	.45	.98	.98	1.00	-.73	-.01	.45	-.02	.45	.96	.96	1.00	-.67
DIF Omit	.01	.02	.00	.02	-.77	-.77	-.73	.99	.01	.01	.01	.01	-.71	-.71	-.69	.98
Hispanic/White																
DSTD-P	-.00	.02	-.00	.02	1.00	1.00	.99	-.52	.00	.02	.00	.02	1.00	1.00	.99	-.42
DSTD-FS	-.01	.02	-.01	.02	1.00	1.00	.99	-.52	.00	.02	.00	.02	1.00	1.00	.99	-.41
MH D DIF	-.08	.27	-.07	.27	.99	.99	1.00	-.50	.01	.28	.02	.29	.99	.99	1.00	-.43
DIF Omit	.00	.01	.00	.01	-.47	-.47	-.47	.96	.00	.01	.00	.01	-.44	-.43	-.46	.96

[1]The speeded criterion is based on all 60 items across both SAT Math sections. The unspeeded criterion is based on 50 items: 20 items in section 1 and 30 items in section 2.

[2]Only data for unspeeded items were included in these calculations (section 1: 20 items and section 2: 30 items).

[3]In the correlation matrix, DIF and DIF Omit variables based on the speeded criterion are presented in the rows, and those based on the unspeeded criterion are in the columns. All correlations are across-criteria correlations.

Table 3A

Descriptive Statistics for Item Difficulty, Differential Item Functioning, and Differential Omission Rates for Verbal Sections of a 1984 SAT[1]

	Section 1[2]						Section 2[2]					
			Correlations						Correlations			
	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF
Female/Male												
Difficulty (PI)	.53	.25					.56	.25				
DSTD-P	.00	.04	-.19				.01	.05	.10			
DSTD-FS	.01	.05	-.19	1.00			.01	.06	.10	1.00		
MH D DIF	.06	.55	-.16	.97	.97		.15	.82	.05	.98	.98	
DIF Omit	-.01	.02	.37	-.69	-.69	-.67	-.01	.02	.14	-.72	-.72	-.64
Black/White												
Difficulty (PI)	.53	.25					.57	.25				
DSTD-P	.01	.03	-.29				-.00	.04	-.39			
DSTD-FS	.01	.04	-.29	1.00			-.00	.05	-.39	1.00		
MH D DIF	.16	.45	-.42	.96	.96		-.06	.59	-.45	.98	.98	
DIF Omit	-.00	.01	.45	-.52	-.52	-.57	-.01	.02	.47	-.51	-.51	-.60
Asian American/White												
Difficulty (PI)	.54	.25					.58	.25				
DSTD-P	-.00	.04	-.24				-.01	.04	-.35			
DSTD-FS	-.00	.05	-.24	1.00			-.01	.04	-.35	1.00		
MH D DIF	-.10	.64	-.36	.94	.94		-.16	.60	-.46	.92	.92	
DIF Omit	.02	.02	-.52	-.52	-.52	-.39	.02	.02	-.36	-.38	-.38	-.25
Hispanic/White												
Difficulty (PI)	.54	.25					.58	.25				
DSTD-P	.01	.03	-.16				.00	.04	-.42			
DSTD-FS	.01	.04	-.16	1.00			.00	.04	-.42	1.00		
MH D DIF	.07	.48	-.25	.97	.97		-.01	.55	-.51	.97	.97	
DIF Omit	-.01	.01	.46	-.60	-.60	-.63	-.01	.02	.61	-.59	-.59	-.64

[1]The speeded total SAT Verbal scaled score (based on 85 items) was used as the matching criterion in the computation of all DIF and Differential Omission statistics.

[2]Only data for unspeeded items were included in these calculations of correlations (section 1: 35 items and section 2: 35 items).

Table 3B

Descriptive Statistics for Item Difficulty, Differential Item Functioning, and Differential Omission Rates for Verbal Sections of a 1985 SAT[1]

	Section 1[2]						Section 2[2]					
			Correlations						Correlations			
	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF
Female/Male												
Difficulty (PI)	.55	.24					.57	.26				
DSTD-P	.01	.04	-.28				-.00	.04	-.07			
DSTD-FS	.01	.05	-.28	1.00			-.00	.04	-.07	1.00		
MH D DIF	.05	.59	-.42	.94	.94		-.07	.74	-.18	.87	.87	
DIF Omit	-.02	.03	.61	-.72	-.72	-.75	-.01	.02	.27	-.66	-.66	-.52
Black/White												
Difficulty (PI)	.56	.24					.58	.26				
DSTD-P	.01	.03	-.22				-.00	.05	-.37			
DSTD-FS	.01	.04	-.22	1.00			-.00	.06	-.37	1.00		
MH D DIF	.13	.46	-.22	.95	.95		.00	.71	-.43	.98	.98	
DIF Omit	-.00	.01	.41	-.46	-.46	-.47	-.00	.02	.57	-.72	-.72	-.71
Asian American/White												
Difficulty (PI)	.57	.24					.59	.26				
DSTD-P	-.00	.03	-.31				-.00	.03	-.36			
DSTD-FS	-.00	.04	-.31	1.00			-.01	.04	-.36	1.00		
MH D DIF	-.04	.47	-.35	.95	.95		-.04	.51	-.33	.96	.96	
DIF Omit	.02	.02	-.54	-.44	-.44	-.36	.02	.02	-.26	-.54	-.54	-.50
Hispanic/White												
Difficulty (PI)	.56	.24					.58	.26				
DSTD-P	.01	.02	-.13				-.01	.03	-.30			
DSTD-FS	.01	.03	-.13	1.00			-.01	.04	-.30	1.00		
MH D DIF	.21	.39	-.02	.89	.89		-.05	.49	-.30	.97	.97	
DIF Omit	-.00	.01	.16	-.57	-.57	-.51	-.00	.01	.37	-.77	.77	-.71

[1]The speeded total SAT Verbal scaled score (based on 85 items) was used as the matching criterion in the computation of all DIF and Differential Omission statistics.

[2]Only data for unspeeded items were included in these calculations of correlations (section 1: 32 items and section 2: 33 items).

Table 4A

Descriptive Statistics for Item Difficulty, Differential Item Functioning, and Differential Omission Rates for Math Sections of a 1984 SAT[1]

	Section 1[2]						Section 2[2]					
			Correlations						Correlations			
	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF
Female/Male												
Difficulty (PI)	.63	.19					.60	.21				
DSTD-P	.00	.04	.28				.00	.03	.25			
DSTD-FS	.00	.05	.28	1.00			.00	.03	.25	1.00		
MH D DIF	.03	.62	.37	.98	.98		.08	.37	.27	.98	.98	
DIF Omit	.01	.02	-.54	-.83	-.83	-.89	.00	.01	-.39	-.57	-.56	-.57
Black/White												
Difficulty (PI)	.64	.19					.61	.21				
DSTD-P	-.01	.04	-.23				.00	.03	-.11			
DSTD-FS	-.01	.05	-.23	1.00			.00	.04	-.11	1.00		
MH D DIF	-.14	.50	-.16	.98	.98		.03	.38	-.11	.98	.98	
DIF Omit	-.00	.01	.26	-.57	-.57	-.60	.00	.01	-.32	-.56	-.55	-.54
Asian American/White												
Difficulty (PI)	.65	.19					.62	.21				
DSTD-P	-.01	.05	-.28				.01	.03	-.03			
DSTD-FS	-.01	.05	-.28	1.00			.01	.04	-.04	1.00		
MH D DIF	-.17	.80	-.26	.96	.96		.12	.51	.09	.96	.96	
DIF Omit	.01	.02	.17	-.78	-.78	-.71	-.00	.01	.01	-.69	-.69	-.66
Hispanic/White												
Difficulty (PI)	.65	.19					.62	.21				
DSTD-P	-.01	.02	-.36				.00	.02	-.10			
DSTD-FS	-.01	.03	-.36	1.00			.00	.02	-.11	1.00		
MH D DIF	-.09	.38	-.37	.99	.99		.04	.27	-.12	.99	.99	
DIF Omit	-.00	.01	.38	-.71	-.71	-.69	-.00	.01	-.06	-.33	-.33	-.37

[1]The speeded total SAT Math scaled score (based on 60 items) was used as the matching criterion in the computation of all DIF and Differential Omission statistics.

[2]Only data for unspeeded items were included in these calculations of correlations (section 1: 20 items and Section 2: 30 items).

Table 4B

Descriptive Statistics for Item Difficulty, Differential Item Functioning, and Differential Omission Rates for Math Sections of a 1985 SAT[1]

	Section 1[2]						Section 2[2]					
			Correlations						Correlations			
	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF	Mean	SD	Diff. (PI)	DSTD-P	DSTD-FS	MH D DIF
Female/Male												
Difficulty (PI)	.65	.21					.59	.21				
DSTD-P	-.00	.04	.14				-.00	.04	.01			
DSTD-FS	-.00	.04	.14	1.00			-.00	.05	.01	1.00		
MH D DIF	-.02	.57	.11	.97	.97		-.01	.58	.08	.98	.98	
DIF Omit	.01	.02	-.36	-.63	-.63	-.59	.01	.02	-.23	-.77	-.77	-.75
Black/White												
Difficulty (PI)	.65	.21					.59	.21				
DSTD-P	-.01	.04	-.32				.00	.03	-.28			
DSTD-FS	-.01	.05	-.32	1.00			.00	.04	-.28	1.00		
MH D DIF	-.14	.54	-.34	.99	.99		.05	.44	-.32	.99	.99	
DIF Omit	.00	.02	.09	-.72	-.72	-.71	.00	.01	.20	-.46	-.46	-.50
Asian American/White												
Difficulty (PI)	.66	.21					.61	.21				
DSTD-P	.00	.03	-.05				-.00	.03	-.20			
DSTD-FS	.00	.03	-.05	1.00			-.00	.03	-.20	1.00		
MH D DIF	.02	.45	-.01	.98	.98		-.01	.45	-.16	.96	.96	
DIF Omit	.01	.02	-.06	-.77	-.77	-.73	.01	.01	-.20	-.70	-.70	-.69
Hispanic/White												
Difficulty (PI)	.66	.21					.60	.21				
DSTD-P	-.00	.02	-.19				.00	.02	-.28			
DSTD-FS	-.01	.02	-.19	1.00			.00	.02	-.28	1.00		
MH D DIF	-.08	.27	-.20	.99	.99		.01	.28	-.28	.99	.99	
DIF Omit	.00	.01	-.01	-.46	-.46	-.46	.00	.01	-.09	-.43	-.42	-.44

[1]The speeded total SAT Math scaled score (based on 60 items) was used as the matching criterion in the computation of all DIF and Differential Omission statistics

[2]Only data for unspeeded items were included in these calculations of correlations (section 1: 20 items and section 2: 30 items).

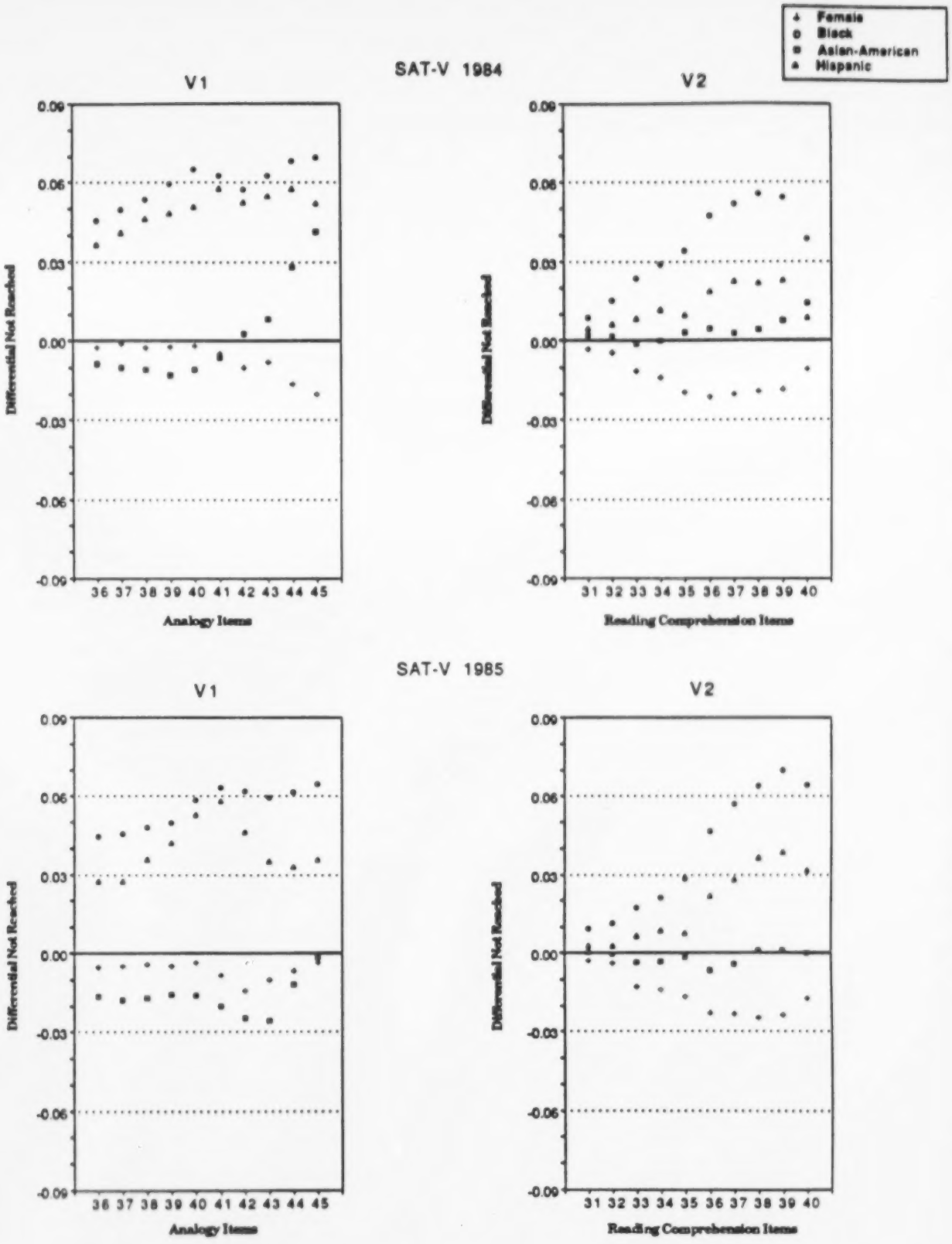

Figure 1. Differential Speededness on the Last Ten Items in the Verbal Sections of a 1984 and a 1985 SAT

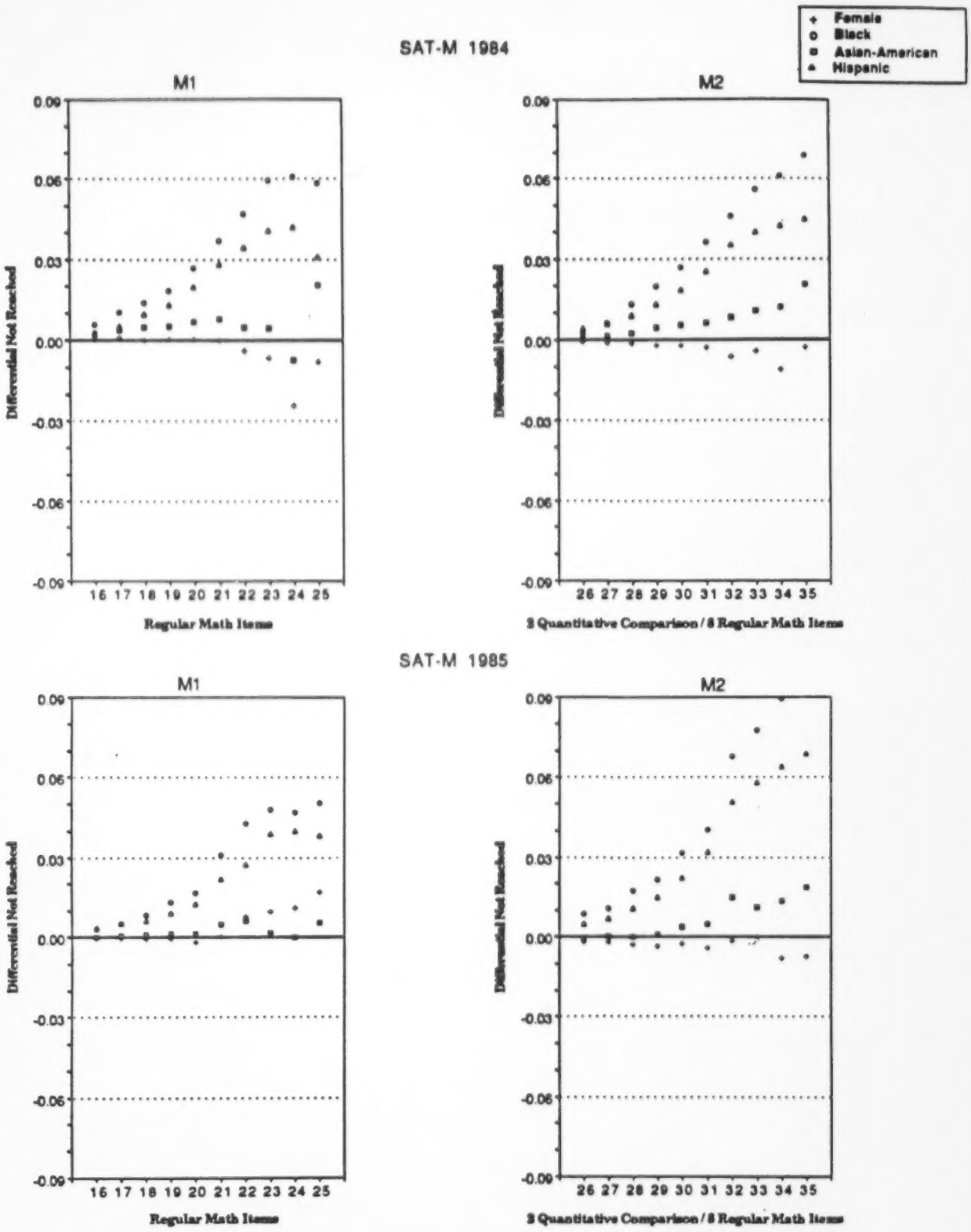

Figure 2. Differential Speededness on the Last Ten Items in the Math Sections of a 1984 and 1985 SAT

Note. The differential not reached value of item 35 for the Black group comparison is 0.11.

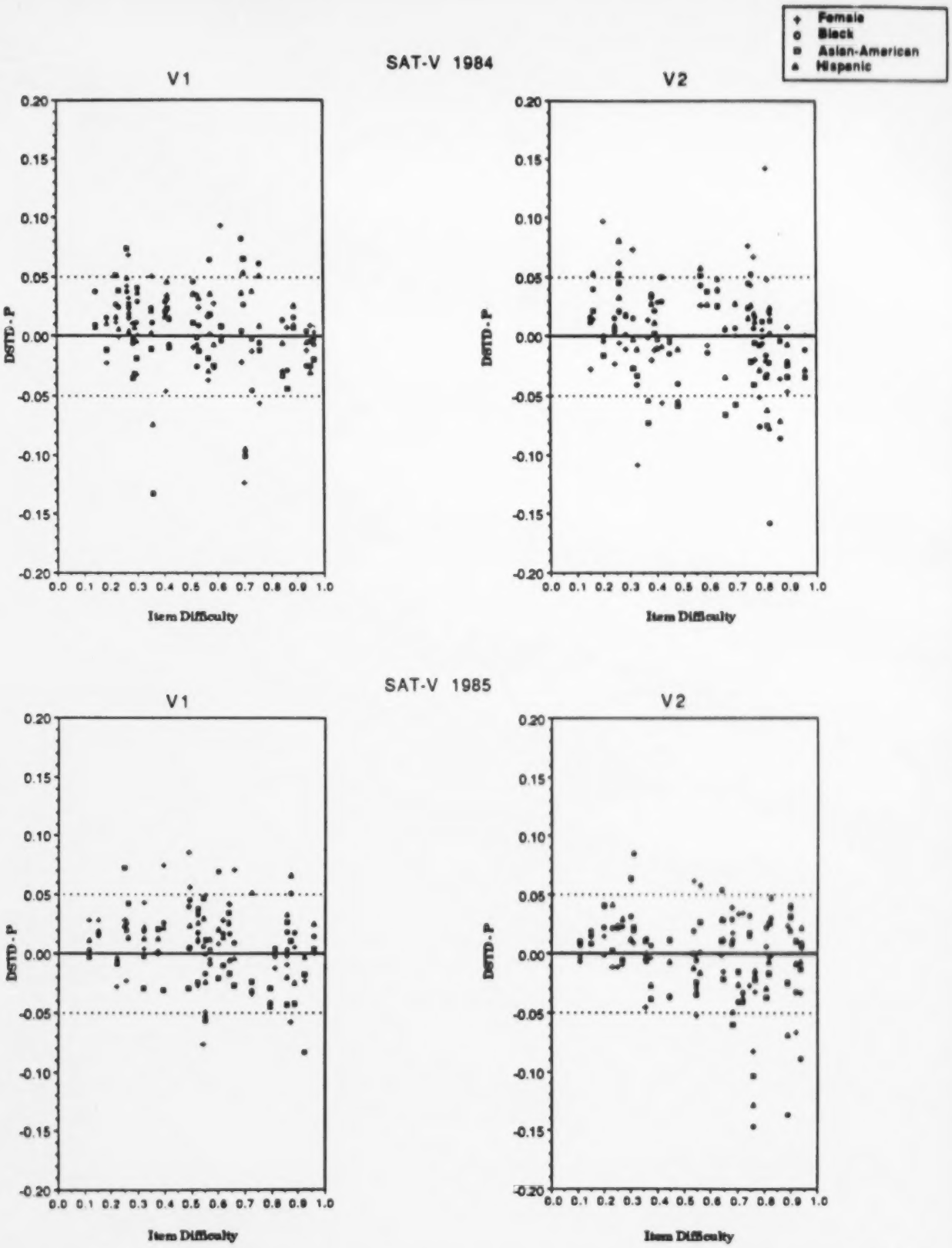

Figure 3. Relationship Between DIF (DSTD-P) and Item Difficulty for the Verbal Sections of a 1984 and 1985 SAT.

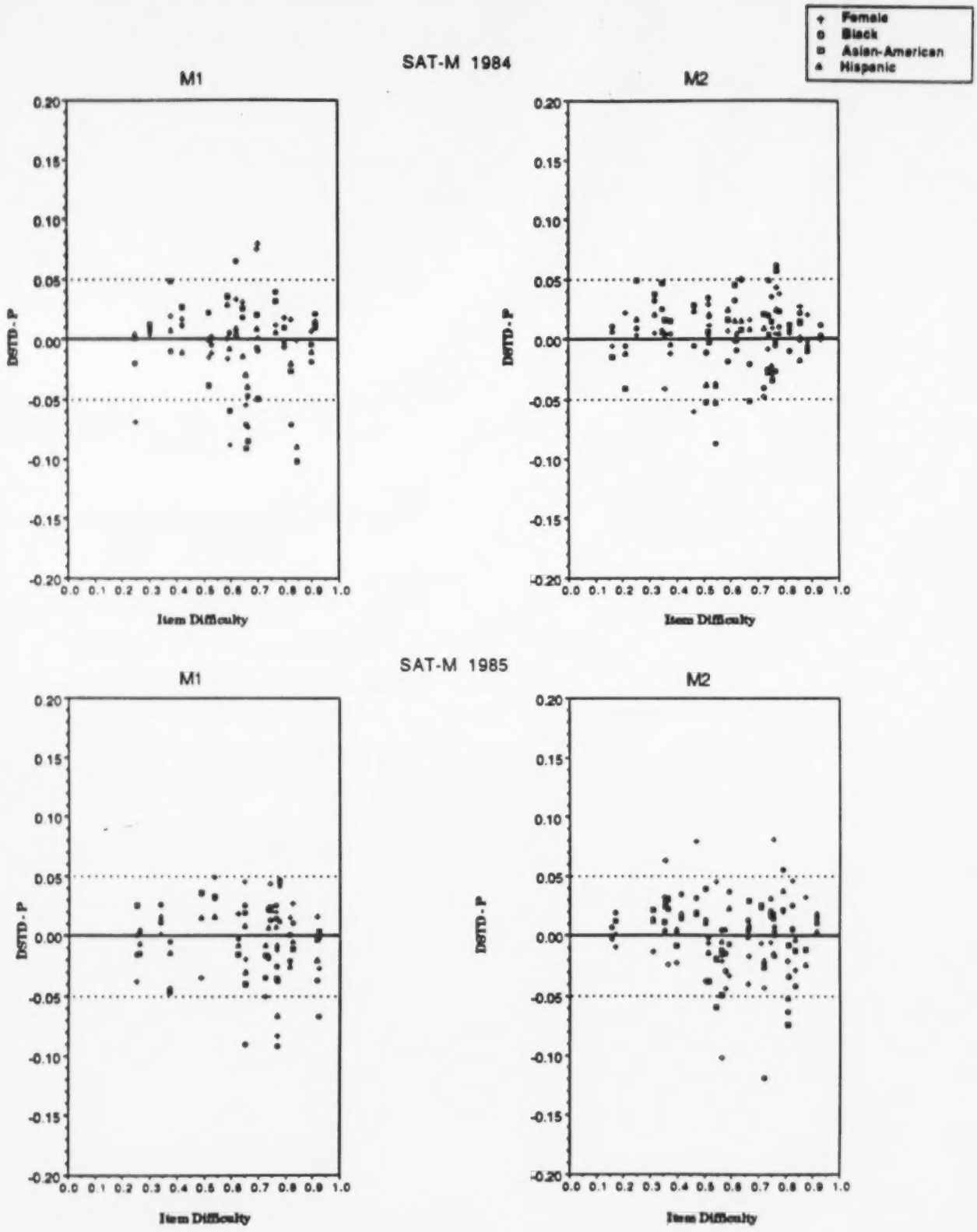

Figure 4. Relationship Between DIF (DSTD-P) and Item Difficulty for the Math Sections of a 1984 and 1985 SAT.

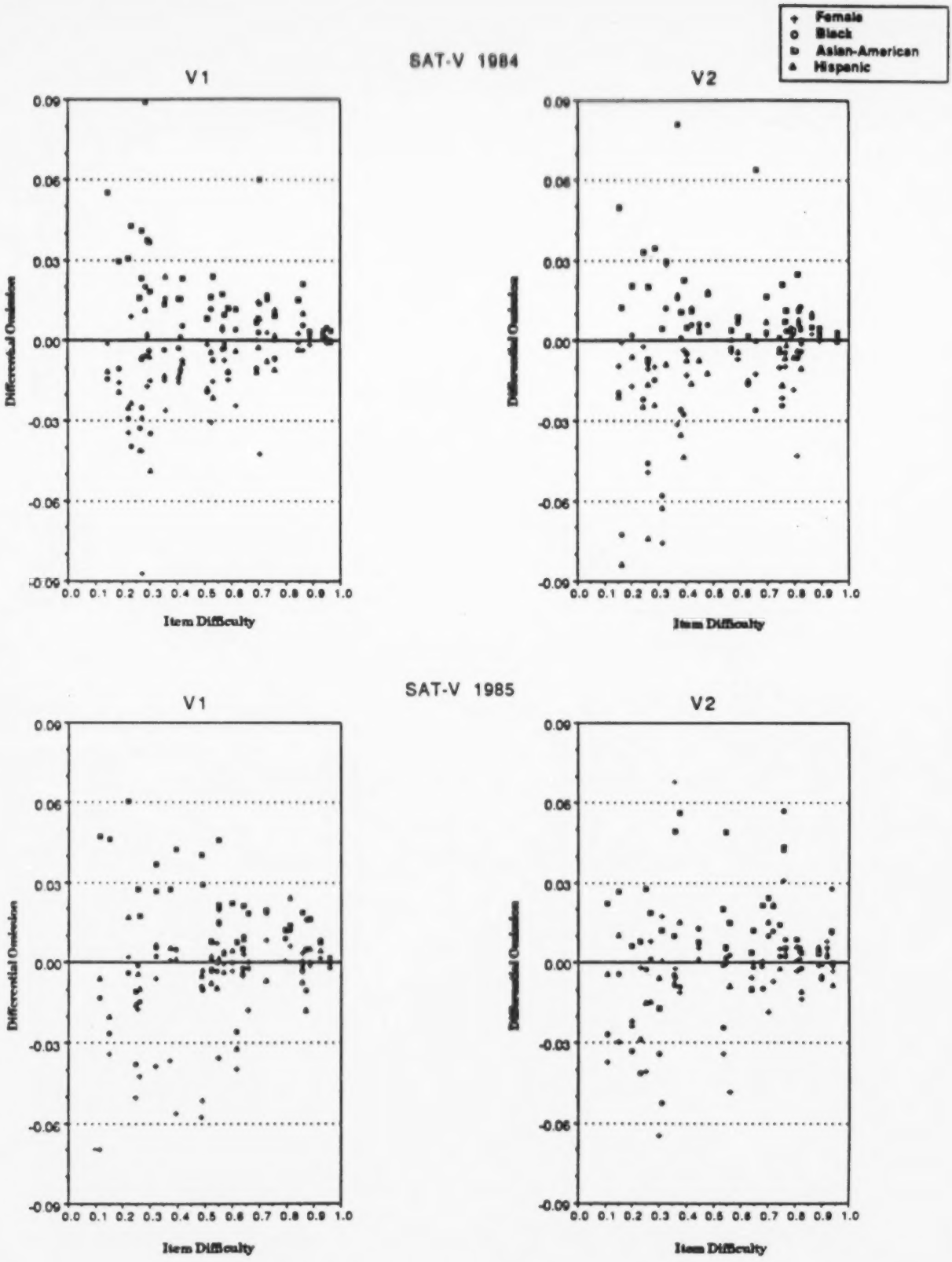

Figure 5. Relationship Between Differential Omission and Item Difficulty for the Verbal Sections of a 1984 and 1985 SAT.

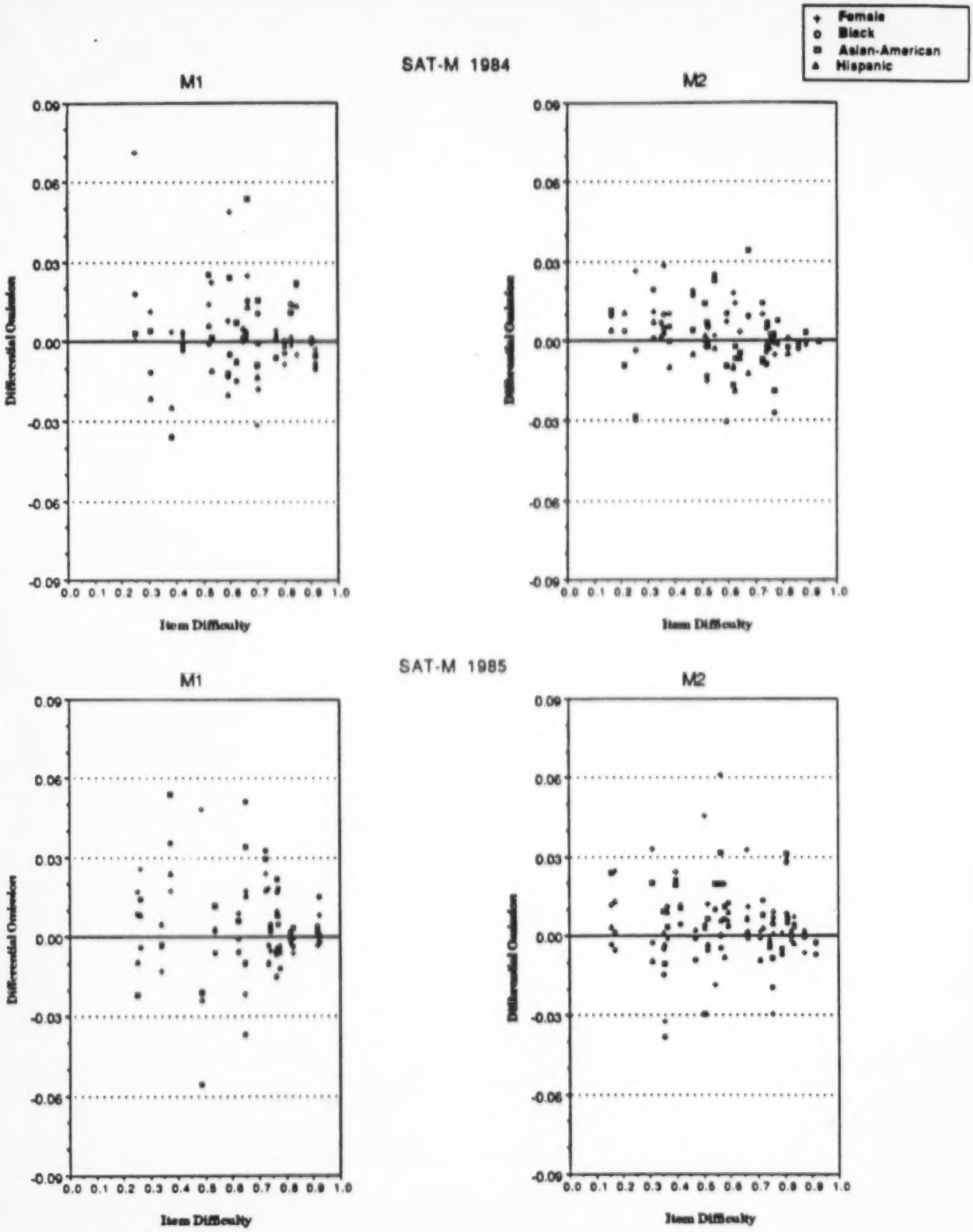

Figure 6. Relationship Between Differential Omission and Item Difficulty for the Math Sections of a 1984 and 1985 SAT.

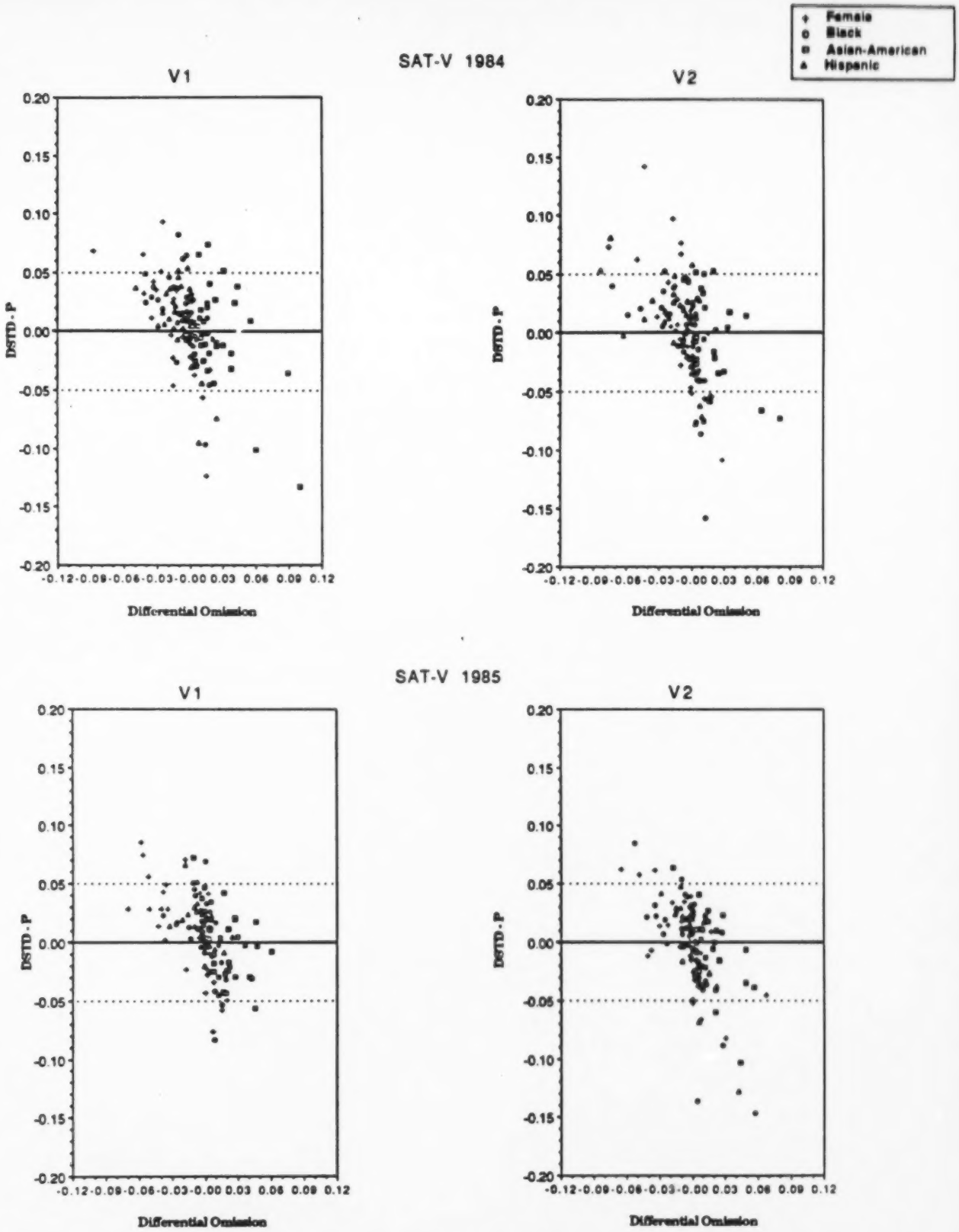

Figure 7. Relationship Between Differential Omission and DIF (DSTD-P) for the Verbal Sections of a 1984 and 1985 SAT.

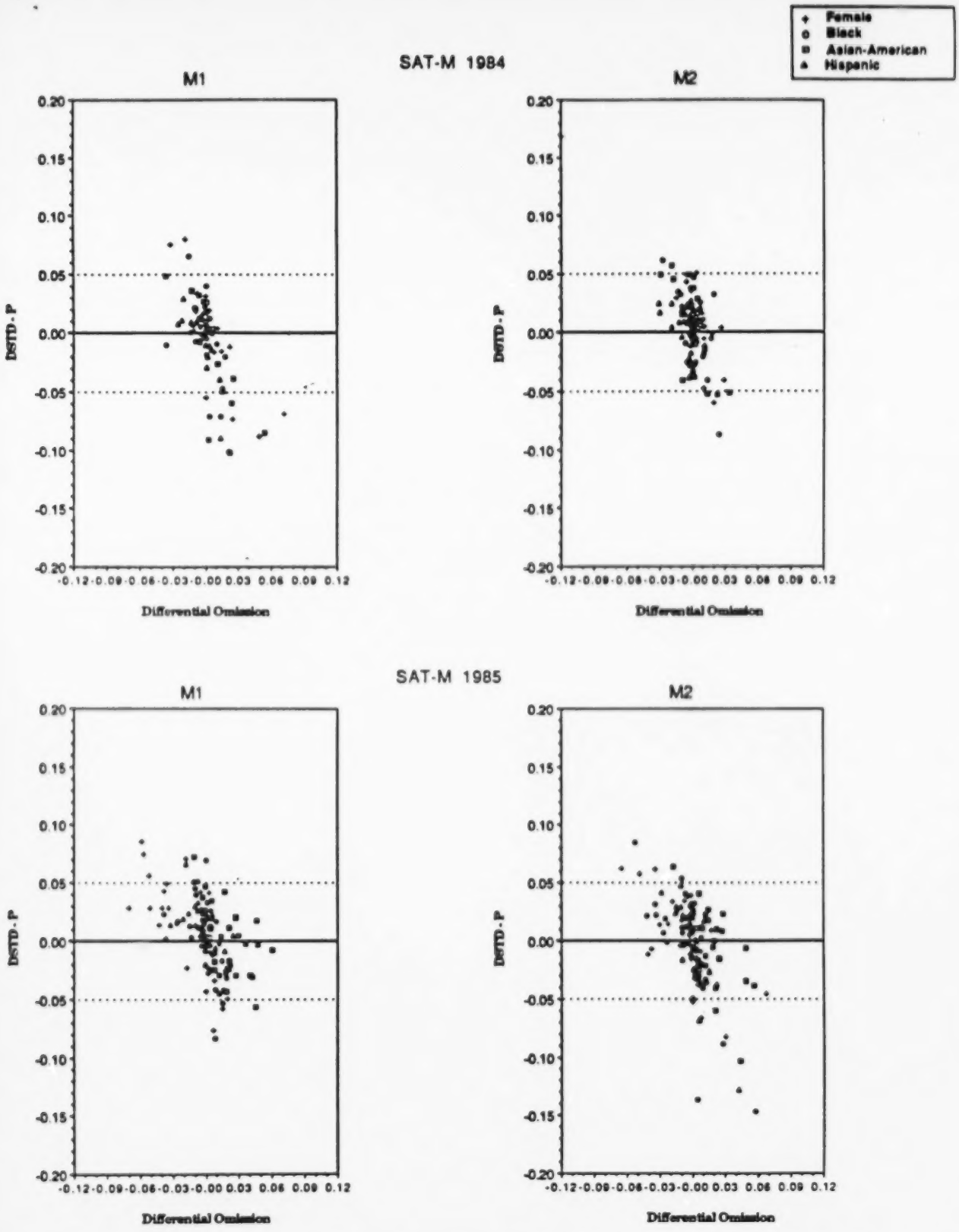

Figure 8. Relationship Between Differential Omission and DIF (DSTD-P) for the Math Sections of a 1984 and 1985 SAT.